

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Open Access 2007 - 2017: Country and University Level Perspective

### Conference or Workshop Item

#### How to cite:

Gyawali, Bikash; Pontika, Nancy and Knoth, Petr (2020). Open Access 2007 - 2017: Country and University Level Perspective. In: In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020 (JCDL '20).

For guidance on citations see [FAQs](#).

© 2020 ACM



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1145/3383583.3398606>

<http://2020.jcdl.org/>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

# Open Access 2007 – 2017: Country and University Level Perspective

Bikash Gyawali  
bikash.gyawali@open.ac.uk  
The Open University  
Milton Keynes, United Kingdom

Nancy Pontika  
nancy.pontika@open.ac.uk  
The Open University  
Milton Keynes, United Kingdom

Petr Knoth  
petr.knoth@open.ac.uk  
The Open University  
Milton Keynes, United Kingdom

## ABSTRACT

Each year the number of Open Access (OA) papers is gradually increasing. We carried out a study investigating 400 universities from 8 countries to examine: i) the total number of OA papers per country, ii) proportion of OA papers published by representative universities in each country classified into three tiers of research quality: high, middle and low, iii) how universities within the same country compare to each other and iv) the growth of OA papers in countries per year. We conclude that among the analysed countries the UK and USA rank first and second respectively, while Russia and India are positioned towards the bottom of the list. We observe no link between the proportion of OA papers published by authors at a university and the university ranking, with some universities in the middle university rank tier having a larger proportion of OA papers than those in the high tier.

## CCS CONCEPTS

• Information systems → Data mining; Digital libraries and archives; • Applied computing → Publishing.

## KEYWORDS

Open Access, Scholarly Data, Data Mining, Growth, Countries, Universities, 2007 – 2017

### ACM Reference Format:

Bikash Gyawali, Nancy Pontika, and Petr Knoth. 2020. Open Access 2007 – 2017: Country and University Level Perspective. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020 (JCDL '20), August 1–5, 2020, Virtual Event, China*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3383583.3398606>

## 1 INTRODUCTION

The impact of the world wide web has affected scholarly communications; journals abandoned the printed version of their volumes and issues and started publishing in an online form. Authors did not need to solely depend on a structured publishing system but were able to disseminate their work more freely via alternative routes of publication. They were also in position to choose the level of use and reuse of their articles for their readers, enabling not only

technically, but also in terms of rights, the furthest possible dissemination of the results of their research. OA had a preeminent role in this swift in scholarly communications, since for its application is required for a paper to be in a digital form, be accessible online and free of cost to the end user. At the same time an author can decide whether to apply limited or no licensing restrictions [19].

OA was defined with the Budapest Open Access Initiative [1] where in addition to the the aforementioned requirements, the routes to OA were also determined. Thus, OA can be delivered either via a publication in an open access journal or via self-archiving in a repository. In OA jargon these are called Gold OA and Green OA respectively. It has been eighteen years since the BOAI and during this time a great amount of OA papers have been published.

This paper analyses the proportion and growth of OA in the years 2007 - 2017 at a country and university level. Although there is previous research studying OA growth [12], [20], [13] and using a variety of databases to count the OA percentage [14], [4], [18], [2], no previous research has, to the best of our knowledge, neither investigated OA growth at university and/or country level nor based their analysis on the same data as ours (more information about our dataset in Section 3). While it is well understood that the proportion of OA articles has been growing, it is important to analyse which institutions, nations (and also OA policies) are most contributing to this growth. This is where our contribution lies.

## 2 RELATED WORK

OA growth can be seen in the numbers reported by the main OA infrastructure services. For example, CORE<sup>1</sup> reports 15 million full texts and 115 million metadata records from 9,775 content providers. BASE<sup>2</sup> has 157 million metadata records from 7,731 providers, while OpenDOAR<sup>3</sup> registry lists 5,301 repositories. The Directory of Open Access Journals<sup>4</sup> lists 14,179 journals and 4,555,376 articles from 130 countries, while the Directory of Open Access Books<sup>5</sup> has 26,308 academic peer reviewed books from 368 publishers. Finally, the ROARMap<sup>6</sup> registry lists 1019 OA policies. According to a report [15] analysing the growth of OA, more than half of the services studied grew for 10%, while 23% of them grew over 20%.

Previous studies measuring the proportion of OA journals and articles discovered 2.6% of OA journals and 3% OA articles in 2004 [14] and 17% of OA articles and 10% of OA journals in 2013 [18]. For 2017, [9] discovered 12.4% of OA journals and 10.3% of OA articles while [3] found 18.4% and 18.9% of OA journals and OA articles

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

JCDL '20, August 1–5, 2020, Virtual Event, China

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7585-6/20/06...\$15.00

<https://doi.org/10.1145/3383583.3398606>

<sup>1</sup>CORE <https://core.ac.uk/>

<sup>2</sup>BASE <https://www.base-search.net/>

<sup>3</sup>OpenDOAR <https://v2.sherpa.ac.uk/pendoar/>

<sup>4</sup>DOAJ <https://doaj.org/>

<sup>5</sup>Directory of Open Access Books <https://www.doabooks.org/>

<sup>6</sup>ROARMap <https://roarmap.eprints.org/>

**Table 1: Count of universities found in THE WUR and MAG for each analysed country**

	Austria	Brazil	Germany	India	Portugal	Russian Federation	UK	USA
THE WUR count	11	46	48	56	13	39	100	172
MAG count	9	34	34	46	11	21	91	154
% Coverage	81.81%	73.91%	70.83%	82.14%	84.61%	53.84%	91.0%	89.53%

respectively. Other research [2] studied the peer-reviewed OA Gold papers in Europe and in the world to discover that the growth for years 1996 - 2012 was 24% per year. Related studies have also measured the volume of OA with regards to both of its two publication routes, Gold and Green OA. A study with 2006 data revealed a 11.3% Gold OA and 8.1% Green OA [4], while a study for 2008 revealed a proportion of 6.6% Gold OA and 14% for Green OA [5]. For 2012, [20] discovered from a sample that 2.4% was Gold OA while 21.4% was Green OA and for the same year another study found 17% was Gold OA and 5% Green OA [12]. Similar studies measure the growth of content self-archived in repositories, including pre-prints. OA self-archived papers increased significantly between 2005 and 2007 [6], while a recent study shows that gradually more and more pre-prints are self-archived in repositories, over 9,000 papers for years 2016 - 2019 [16]. We distinguish our work from these studies because we:

- present for the first time OA growth analysis at **country** and **university** level.
- use a unique dataset combining CORE, the world’s largest aggregator of OA content, with Microsoft Academic Graph (MAG), the only freely available database of information about scholarly publications.

### 3 METHODOLOGY

In this research, we examined OA performances of eight countries - Austria, Brazil, Germany, India, Portugal, Russia<sup>7</sup>, United Kingdom and United States - and their universities, which are included in the Times of Higher Education World University Rankings 2020 (THE WUR)<sup>8</sup>. In contrast to other similar tools<sup>9, 10</sup>, THE WUR provides a direct indicator of how universities rank with regards to research as well as citations. The countries were selected as an appropriate mix of developed and less developed countries. The universities within each countries were selected based on their availability in MAG and only the papers published in the years 2007–2017 were considered (see Section 3.2). Specifically, we present results for years 2007–2017 in terms of:

- (1) Percentage of OA papers published by representative universities that ranked at high, middle and low tiers in selected countries.
- (2) Total count of OA articles published by each country, and
- (3) Growth of OA content for each country for each year.

### 3.1 Data Preparation

We used MAG [17] as a reference dataset to identify all publications relevant to our study. MAG is a graph of scholarly publications organized into database tables, which record different information related to scholarly publications such as citations, author names, institution (universities as well as other publishing bodies) names and publication years. MAG ranks competitively to alternative resources that index research papers [7] and is freely available for use – this made it an ideal choice for our work. The following MAG tables<sup>11</sup> were relevant to us:

**affiliations** – Contains geographic coordinates (*latitude*, *longitude*) and unique identifiers (*affiliationid*) for various institutions along with their names (*normalisedname*).

**paperauthoraffiliations** – Contains records of all papers (*paperid*) published by all institutions in the database (*affiliationid*).

**papers** – Contains metadata of papers (*paperid*) such as year of publication.

To determine the OA status of a paper we used CORE Discovery, a service that finds links to freely accessible copies of research papers using CORE [11] as well as multiple external services. It delivers state-of-the-art performance compared to other existing discovery tools in terms of both content coverage and precision [10].

### 3.2 Data

We used geographic coordinates to identify countries<sup>12</sup> for institutions in MAG. For each country in our study, we matched the universities with the institutions identified in MAG for the corresponding country. The matching was done based on the institution name – we normalised<sup>13</sup> the names of our universities and checked for string match with the *normalisedname* of institutions in MAG. Table 1 lists the count of universities in THE WUR and the corresponding matches we found in MAG using this approach. This accounts for a total of 400 universities included in our study. The lowest coverage was observed for Russia (53.84%) while others had a fairly good coverage with UK having the highest (91.0%). Subsequently, we extracted lists of unique papers along with their year of publication for each university in our study using the MAG tables mentioned above. Finally, we enriched this dataset with OA status information, true/false, for each paper in our collection as was given by CORE Discovery. We note the following:

- Our dataset compiles OA information at the granularity of papers published by individual universities. This enables for a straight-forward study of university level OA analysis, which was, to our knowledge, not previously conducted.

<sup>7</sup>For brevity reasons, we use Russia instead of Russian Federation.

<sup>8</sup><https://tinyurl.com/y2n8kbv6>

<sup>9</sup>ARWU (<http://www.shanghairanking.com/>)

<sup>10</sup>QS world ranking (<https://tinyurl.com/y46u4un8>)

<sup>11</sup>The complete MAG data schema is available at <https://tinyurl.com/v4r5tfv>

<sup>12</sup>based on <https://tinyurl.com/v2gypq9>

<sup>13</sup>lowercase, punctuation removal and ASCIIification

- The same paper can belong to multiple authors affiliated with the same university. We considered such instances to be **duplicates** and preserved only a single instance of the paper for that university. Also, a paper can have multiple authors affiliated with different universities. In such cases, we included a single instance of the paper for each of the universities concerned.
- To perform the OA study at a country level, we merged the records from all universities in the corresponding countries after eliminating possible **duplicates**, i.e. the records of papers with multiple authors from different universities within the same country were preserved only once.
- Our dataset is based on the 2018 release of MAG and CORE Discovery. In fairness to duration needed for deposit of OA content into repositories and harvesting from them, we extracted and analyzed papers published until 2017.

In order to perform university level OA analysis, we further identified **representative** universities for each country in our study. Within each country, we first classified the universities into three tiers – “high”, “middle” and “low” based on their ranks by “research” score in THE WUR. From each tier  $x$  of each country  $y$ , we selected five<sup>14</sup> universities with median rankings within that tier and designated them as representative universities of the tier  $x$  for the corresponding country  $y$ . In total, this resulted to 108 representative universities for our university level OA analysis.

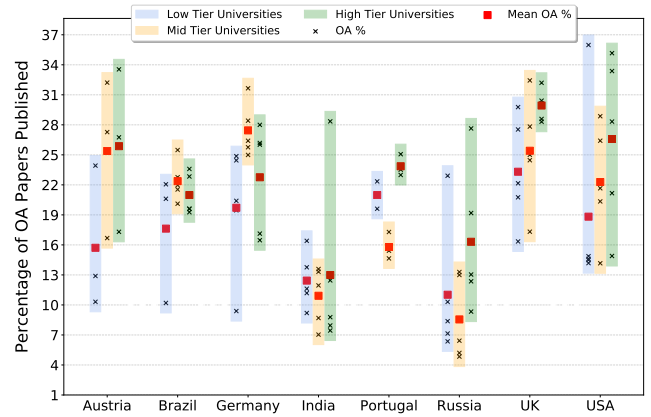
## 4 RESULTS

Figure 1 shows the percentage of OA publication by the representative universities for each country for the years 2007 - 2017. With regards to the universities tiers, it seems that the data do not present a specific pattern. In Austria, the high and middle tier universities have a close percentage of OA papers (33.54% and 32.21%) with the mean for these two tiers being 25.86% and 25.39% respectively. It is interesting that in Germany the middle tier ranks higher, with a mean of 27.44%, comparing to the high tier with 22.75%. Portugal is the country with the lowest data variation, but we observe that the middle tier universities perform poorly, mean OA 15.78%, compared to the universities in the low tier, mean OA 20.97%. The UK has the highest means: high - 29.93%, middle - 25.40% and low - 23.13%, while India has the lowest accumulative mean 11.35% with Russia ranking slightly higher 11.72%.

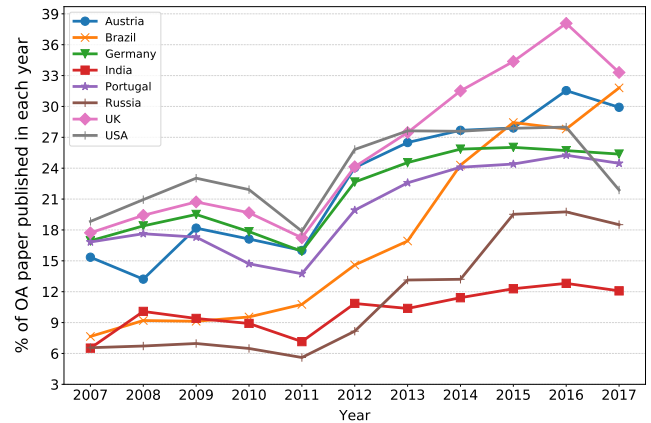
**Table 2: Count and percentage of OA papers per country**

	% OA	Count OA	Count Unknown	Total
Austria	23.22%	27,928	92,345	120,273
Brazil	19.54%	117,405	483,227	600,632
Germany	22.27%	133,805	466,992	600,797
India	10.73%	27,918	232,245	260,163
Portugal	21.16%	26,850	100,010	126,860
Russia	14.49%	13,700	80,834	94,534
UK	26.68%	339,410	932,428	1,271,838
USA	24.04%	938,287	2,963,252	3,901,539

<sup>14</sup>three in case of Austria and Portugal due to low number of total universities



**Figure 1: Representative universities with %OA papers**



**Figure 2: Percentage of OA papers published in each year**

We also estimated the OA count for all the universities in all the eight studied countries. In Table 2, we see that the UK has the highest percentage of OA outputs, 26.68%. This percentage is higher than the previously mentioned studies in the related work, but this difference should be expected because in the meantime, OA growth is expected to have increased. Second and third are USA (24.04%) and Austria (23.22%) while Germany (22.27%) and Portugal (21.26%) rank lower with a small difference of 1%.

Figure 2 shows the percentage of OA papers published for each year per country. The results present an expected pattern; earlier years, 2007 - 2011, show low numbers of OA papers with a steady growth, but numbers increase significantly after 2012. In 2016 there is a peak for both Austria and the UK with slight drop in 2017. This could probably be caused by difficulties with aggregating recent data from providers and/or OA discovery issues. Although our analysis shows Brazil has had lower numbers comparing to other countries, it's percentage of OA papers is drastically increasing, while Germany and Portugal follow a steady route.

**Table 3: Funder and Institutional OA Policies per Country**

	Institutional	Funder
Austria	13	1
Brazil	22	1
Germany	71	0
India	9	1
Portugal	25	1
Russia	4	0
UK	19	93
USA	78	4

## 5 DISCUSSION

OA growth is globally promoted via international funder policies, with most recent being for example CoalitionS' PlanS<sup>15</sup>, and actions such as the Open Access 2020 Initiative<sup>16</sup>. For the countries we studied, ROARMap lists the number of policies (Table 3), showing how many funder and institutional OA policies have been established in each country. The UK has the greatest number of both funder (19) and institutional policies (93), while USA comes second in the number of institutional OA policies (78), and third in the number of funder policies (4). The rest of the countries hold the second place in terms of funder policies (1). In Section 4, we see that UK ranks first in its proportion of OA papers, while India ranks last among the studied countries. Portugal, with only one funder policy, but with 25 institutional policies has an OA percentage of 21.16%. Another example from the tested countries, Brazil, has a fairly low number of funder (1) and institutional policies (22) - especially when compared with the number of universities in the country - and its OA percentage is slightly lower than Portugal's.

A recent study [8] discovered that the UK is a leader not only in the growth of OA but also leads in the shortest time lag, by observing that strict time limited deposit requirements increase OA counts. If we take into account the number of funders and institutional OA policies, they seem to contribute to the growth of OA in the country. It will be useful if future work could quantify the extent to which the choice of making a paper OA is driven by a national or an institutional policy vs authors own desire, providing a strong evidence for adoption of these policies in other countries.

This study does not make a distinction between OA papers published in journals and submitted in repositories, but takes into consideration this count as a whole. It also did not attempt to study the factors that could promote an increase in the percentage of OA papers and whether OA funder policies are a decision-making factor. Such studies could be investigated in future work.

## 6 CONCLUSION

This study investigated the OA scenario for eight selected countries and their universities; the latter chosen based on their rankings in THE WUR. The OA count and proportions were calculated based on data from CORE, the world's largest aggregator of OA papers, and MAG, the largest freely available dataset of scholarly publications.

We found that as a country the UK has the highest proportion of OA papers; the UK universities produce the largest amount of OA papers and have the largest mean. The rest of the order at a country level is as follows: USA, Austria, Germany, Portugal, Brazil, Russia and India. We could not find a significant pattern with regards to the percentage of OA papers from the universities in our sample and we also found that some middle tier universities rank higher in the percentage of OA papers than the high tier universities in their own country. The key message of our paper is that OA is ever so growing and we are observing more universities choosing to publish and/or disseminate research papers using the OA routes.

## 7 ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under Grant Agreement no 824612 related to the Observing and Negating Matthew Effects in Responsible Research and Innovation Transition (ON-MERRIT) project.

## REFERENCES

- [1] [n.d.]. *Budapest Open Access Initiative*. Retrieved January 14, 2020 from <https://www.budapestopenaccessinitiative.org/>
- [2] E. Archambault, D. Amyot, P. Deschamps, A. Nicol, L. Rebout, and G. Roberge. 2013. *Proportion of Open Access Peer-Reviewed Papers at the European and World Levels 2004 - 2011*. Technical Report. Science-Metrix Inc. for the European Commission DG Research and Innovation, Brussels.
- [3] B-C. Björk and T. Kormeakaki. 2019. Adoption of the open access business model in scientific journal publishing - A cross-disciplinary study. *College and Research Libraries* (2019).
- [4] B-C. Björk, A. Roos, and M. Lauri. 2009. Scientific journal publishing: yearly volume and open access availability. *Information Research* 14, e391 (2009).
- [5] B-C. Björk, P. Welling, M. Laakso, P. Majlender, T. Hedlund, and G. Gudnason. 2010. Open access to the scientific journal literature: situation 2009. *PLoS ONE* 5 (2010).
- [6] S. Harnad, T. Brody, F. Vallières, L. Carr, S. Hitchcock, Y. Gingras, C. Oppenheim, C. Hejrn, and E. Hilf. 2008. The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update. *Serials Review* 34, 1 (2008), 36-40.
- [7] Anne-Wil Harzing and Satu Alakangas. 2017. Microsoft Academic: is the phoenix getting wings? *Scientometrics* 110, 1 (2017), 371-383.
- [8] D. Herrmannova, N. Pontika, and P. Knoth. 2019. Do Authors Deposit on Time? Tracking Open Access Policy Compliance. In *2019 ACM/IEEE Joint Conference on Digital Libraries*.
- [9] M. Jubb, A. Plume, S. Oeben, L. Brammer, R. Johnson, and C. Bütün. [n.d.]. *Monitoring the Transition to Open Access - December 2017*. Retrieved January 14, 2020 from <https://tinyurl.com/yb8vk4j6>
- [10] Petr Knoth and Matteo Cancellieri. 2019. *Analysing the performance of open access papers discovery tools*. <https://tinyurl.com/ycr67l8f>
- [11] Petr Knoth and Zdenek Zdrahal. 2012. CORE: three access levels to underpin open access. *D-Lib Magazine* 18, 11/12 (2012). <http://oro.open.ac.uk/35755/>
- [12] M. Laakso and B-C. Björk. 2012. Anatomy of open access publishing: a study of longitudinal development and internal structure. *BMC Med* 10, 124 (2012).
- [13] M. Laakso, P. Welling, H. Bukvova, L. Nyman, Björk, and T. Hedlund. 2011. The development of open access journal publishing from 1993 to 2009. *PLoS One* 6, e20961 (2011). <https://doi.org/10.1371/journal.pone.0020961>
- [14] C. McVeign. [n.d.]. *Open Access Journals in the ISI Citation Databases: Analysis of the Impact Factors and Citation Patterns - A Citation Study from Thompson Scientific*. Retrieved January 14, 2020 from <https://tinyurl.com/ybnorw74>
- [15] H. Morrison. [n.d.]. *Dramatic Growth of Open Access 2019*. Retrieved January 14, 2020 from <https://tinyurl.com/y8nr378k>
- [16] T. Narock and E. Goldstein. 2019. Quantifying the growth of preprint services hosted by the Center for Open Science. *Publications* 7, 2 (2019), 44.
- [17] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-june Paul Hsu, and Kuansan Wang. 2015. An overview of microsoft academic service (mas) and applications. In *Proceedings of the 24th international conference on world wide web*. ACM, 243-246.
- [18] D. Solomon, M. Laakso, and B-C. Björk. 2013. A longitudinal comparison of citation rates and growth among open access journals. *Journal of Informatics* 7, 3 (2013), 642-650. <https://doi.org/10.1016/j.joi.2013.03.008>
- [19] P. Suber. 2012. *Open Access*. MIT, Cambridge, Massachusetts.
- [20] V. Larivière, Y. Gargouri, Y. Gingras, L. Carr, and S. Harnad. [n.d.]. *Green and gold open access percentages and growth, by discipline*. <https://arxiv.org/abs/1206.3664>

<sup>15</sup>PlanS <https://www.coalition-s.org/>

<sup>16</sup>OA2020 <https://oa2020.org/progress-report-nov2019/>